The effects of elaboration on working memory and long-term memory across age

Bartsch, Lea M ; Oberauer, Klaus

Abstract: Free time to attend to and process information in working memory is key in promoting immediate and delayed retention. One candidate process to cause this benefit is elaboration. We conducted three experiments with young adults – two of which included older adults – to investigate whether free time is used for elaboration, and whether elaboration causes the free-time benefit. Participants remembered lists of nouns, interleaved with short or long free-time intervals, or with filler words connecting all the nouns into a meaningful sentence to assist elaboration. For young adults, assisted elaboration through sentences, and the additional instruction to form a mental image, benefited performance in a working-memory test as much as longer free time, but not more. In contrast, for a delayed test of long-term memory, the benefits of sentence elaboration exceeded those of longer free time. Older adults did not benefit from assisted elaborations in the delayed test, providing further evidence that the long-term memory deficit of older adults arises at least in part from a deficit in elaboration. This elaboration deficit is not driven by a deficit in generating richer representations.

DOI: https://doi.org/10.1016/j.jml.2020.104215

Originally published at:
Bartsch, Lea M; Oberauer, Klaus (2021). The effects of elaboration on working memory and long-term memory across age. Journal of Memory Language, 118:104215.
DOI: https://doi.org/10.1016/j.jml.2020.104215

Journal of Memory and Language 118 (2021) 104215

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The effects of elaboration on working memory and long-term memory across age☆

Lea M. Bartsch *, Klaus Oberauer

University of Zurich, University Research Priority Program "Dynamics of Healthy Aging", Switzerland

ARTICLE INFO

Keywords:
Elaboration
Mental imagery
Working memory
Long-term memory
Aging

ABSTRACT

Free time to attend to and process information in working memory is key in promoting immediate and delayed retention. One candidate process to cause this benefit is elaboration. We conducted three experiments with young adults – two of which included older adults – to investigate whether free time is used for elaboration, and whether elaboration causes the free-time benefit. Participants remembered lists of nouns, interleaved with short or long free-time intervals, or with filler words connecting all the nouns into a meaningful sentence to assist elaboration. For young adults, assisted elaboration through sentences, and the additional instruction to form a mental image, benefited performance in a working-memory test as much as longer free time, but not more. In contrast, for a delayed test of long-term memory, the benefits of sentence elaboration exceeded those of longer free time. Older adults did not benefit from assisted elaborations in the delayed test, providing further evidence that the long-term memory deficit of older adults arises at least in part from a deficit in elaboration. This elaboration deficit is not driven by a deficit in generating richer representations.

Introduction

In theories of human memory, a distinction is often made between working memory (WM) and long-term memory (LTM). WM is understood as a system that holds mental representations temporarily available for processing, with limited capacity. In LTM information is stored more permanently with presumably unlimited capacity (Cowan, 2008). What is thought to be common to both memory systems is the central role of *control processes* (Atkinson & Shiffrin, 1968). Control processes refer to those that control what is retained in WM and in LTM. Researchers have tried to isolate these processes and investigate their underlying mechanisms as well as their benefits for both WM and LTM. Furthermore, control processes have been proposed to undergo changes throughout the lifespan, and have been hypothesized to be at least partly responsible for age-related declines in memory functioning (Bartsch, Loaiza, Jäncke, Oberauer, & Lewis-Peacock, 2019; Dunlosky & Hertzog, 1998; Loaiza & McCabe, 2013; Shing et al., 2010).

One of these processes is called *elaboration,* and it is understood as enriching the memory representation of an item by activating many aspects of its meaning and by linking it into the pre-existing network of semantic associations (Craik & Tulving, 1975; Greene, 1987). This integration is assumed to lead to a richer memory trace which then is more easily reactivated during recall (Galli, 2014). Other studies have shown that *distinctiveness* drives the beneficial effects of elaboration on memory. For instance, well elaborated items are less likely to be confused with similar stimuli (Gallo, Meadow, Johnson, & Foster, 2008), and items that are associable to more features are better remembered than items with less associable features (Hargreaves, Pexman, Johnson, & Zdrazilova, 2012). Taken together, according to the literature to date, the mechanism underlying elaboration is the enrichment of a memory trace with item-specific features. This makes the elaborated memory trace accessible through multiple retrieval cues. In addition, it results in a unique representation that is easily distinguishable because the distinct representation does not compete with other memory traces. In the same vein, Klein and Loftus, (1988) proposed that elaboration benefits memory through the creation of multiple routes for retrieval but also through inference-based reconstruction in case the retrieval effort fails.

Evidence for the beneficial effect of elaboration on long-term memory is two-fold: First, both young and older adults report engaging in

☆ Author Note: We thank Andra Arnicane, Gary Hoppeler and Raffael Schmitt for helping with data collection. The data and the analysis scripts can be accessed in the Open Science Framework (https://osf.io/4n9y3). The project was funded by the University Priority Program Dynamics of Healthy Aging.

* Corresponding author at: Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmuehlestrasse 14/22, 8050 Zurich, Switzerland.
E-mail address: l.bartsch@psychologie.uzh.ch (L.M. Bartsch).

https://doi.org/10.1016/j.jml.2020.104215
Received 13 January 2020; Received in revised form 19 December 2020; Accepted 21 December 2020
Available online 12 January 2021

elaborative strategies, such as imagery or sentence generation, spontaneously (Bailey, Dunlosky, & Hertzog, 2009; Dunlosky & Hertzog, 2001; Richardson, 1998). Second, orienting tasks inducing a richer processing of the memory material lead to better LTMs (Craik & Tulving 1975). Subsequent work has shown that directly instructing people to engage in elaboration improves their episodic long-term memory (e.g., Bartsch, Singmann, & Oberauer, 2018; Davachi, Maril, & Wagner, 2001). In contrast, the role of elaboration for WM is yet to be determined.

Elaboration as a candidate process in promoting WM

Two findings have given rise to the idea that elaboration might help not only LTM but also WM: First, a subset of participants report using elaboration during WM tasks, and those who do tend to perform better (Bailey, Dunlosky, & Kane, 2011). For instance, Bailey and colleagues asked people about their strategies during a complex-span task and found that two forms of elaboration were frequently reported: Forming a sentence to combine the to-be-remembered words (14% of trials) and forming mental images of the memoranda (8% of trials). Together, varieties of elaboration were the most frequently reported strategies after rote repetition and reading (Bailey, Dunlosky, & Hertzog, 2009; see also Dunlosky & Kane, 2007). Furthermore, the degree to which an individual reportedly used these two elaborative strategies was positively correlated with their WM span performance, leading these authors to propose that elaboration is an effective maintenance strategy for WM. However, an alternative explanation for the latter finding is that those participants which have good WM, have more capacity to engage in elaborative strategies. Causality can only be inferred from the direct manipulation of the occurrence of elaboration, so in order to differentiate these two explanations, here, we examine the effect of an experimental manipulation of elaboration on behavior.

The second hint that elaboration could help WM is more indirect: Performance in WM tasks is improved by additional free time under certain conditions: (a) if the time is provided between the to-be-remembered items (e.g. Oberauer & Lewandowsky, 2016; Ricker & Hardman, 2017; Tan & Ward, 2008), and (b) more consistently if the items are presented visually rather than auditorily (e.g. Penney, 1975). To date, it is unclear what is causing this effect, but elaboration is a strong candidate: With more free time between the to-be-remembered information, people could engage in this process more.

Elaboration as a candidate process has moved to the focus of our research, as a recent study showed strong evidence against another control process – rehearsal –causing the beneficial effect of free time. Specifically, in that study the extent of cumulative rehearsal was increased through a cumulative rehearsal instruction (Souza & Oberauer, 2018; Exp. 1) as well as a fixed-rehearsal strategy (Exp. 2), yet there was no benefit for WM recall compared to a free-rehearsal baseline. Furthermore, that study showed that the beneficial effect of free time between individual memory items is also observed when articulatory rehearsal is blocked, specifically in the case of concrete and highly imaginable words, which are in general easier to elaborate (Exp 3). This result suggests that subjects might have engaged in a form of elaboration, but as of now, there is no direct evidence to support this assumption.

First attempts to experimentally induce elaboration to test is effect on WM, have yielded no evidence for it: Instructing participants to form a vivid mental image of parts of the memoranda after list presentation did not benefit WM, although it improved LTM (Bartsch, Loaiza, Jäncke, Oberauer, & Lewis-Peacock, 2019; Bartsch, Singmann, & Oberauer, 2018; Bartsch & Oberauer, 2019). More specifically, in these studies, subjects encoded lists of six words, followed by a processing phase in which they were instructed to re-read, refresh, elaborate by forming a mental image, or to simultaneous refresh and elaborate half of the studied list. Compared to a no-processing baseline neither refreshing nor elaboration did much to improve WM performance.

There might be a reason why the elaboration instruction in our earlier studies did not improve WM: Asking participants to elaborate a set of words after an entire word list has been encoded into WM might make it too hard for them to access the words they should elaborate. Elaboration might be easier when it can occur *in between* presentation of individual items. This assumption receives further plausibility by the fact that free time improves WM only when added in *between* list items, not when added at the end of the list (Oberauer & Lewandowsky, 2016).

The current study was designed to make elaboration as easy as possible for participants. To this end, we provided the enriching information via sentences rather than asking participants to generate the enriched representations themselves. The technique of forming sentences of individual to-be-remembered words has previously been reported in strategy assessment, and used in training studies (McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003).

As some participants in strategy-report studies reported forming images of the memory items, in one condition we added the instruction to form an image of the sentence, with the aim to boost the effectiveness of elaboration further: An early study investigating the effect of sentence generation and mental imagery on long-term memory recognition of paired associates showed that the visualization of a mental image led to better recall than simply being presented with the word pair embedded in a sentence (Bower & Winzenz, 1970). This finding suggests that embedding to-be-remembered words in sentences helps people to elaborate because they can draw on the scene or image described in the sentence rather than having to invent a mental image on the spot. Participants might obtain a larger benefit from elaboration – not only for LTM but also for WM – if, in addition to presenting a sentence, mental imagery of the sentence's meaning is encouraged by the instruction.

In summary, the first goal of the present study is to test the hypothesis that elaboration of information in WM, when facilitated by providing enriching information, improves not only long-term retention but also performance in an immediate test of WM. We do so by integrating the three approaches from past research on elaboration in one paradigm: (1) strategy reports, (2) adding free time in between memory items to enable elaboration, as well as (3) experimentally instructing elaboration.

Age-related shortfalls in elaboration processes relate to LTM deficits

Both WM and LTM decline in old age (McDaniel, Einstein, & Jacoby, 2008; Park et al., 2002; Zacks, Hasher, & Li, 2000). The LTM deficit of older adults can in part be attributed to a deficient effectiveness of elaboration in older compared to young adults (Bartsch, Loaiza, Jäncke, et al., 2019; Bartsch & Oberauer, 2019). More precisely, we have previously shown that older adults did not benefit from elaboration in LTM although their brain activation patterns during the formation of mental images at encoding were differentiable from a repeated reading condition – similar to young adults (Bartsch, Loaiza, Jäncke, et al., 2019).

Smith (1980) has argued that the reason why older adults' LTM does not benefit from elaboration is that they have difficulties generating the necessary enrichment of the learning material. This *generation-deficit hypothesis* states that older adults exhibit smaller elaboration benefits on LTM, compared to young adults, when they have to generate the richer representations themselves. In line with this claim, Rankin and Collins (1985) provided evidence that older adults' memory benefited equivalently to young adults when elaborations were given to them in the form of sentences, but they were less likely than young adults to generate relevant elaborations themselves. Similarly, Cherry, Park, Frieske, and Rowley (1993) showed that explanatory elaborations provided in the form of sentences at encoding enhanced delayed memory for target adjectives in young and older adults, but only when the elaborations were given again at retrieval. Age-related production deficits for encoding strategies have been reported more generally as well: For instance, older adults are less likely than young adults to spontaneously use effective strategies when studying paired associates, but they can successfully use them if instructed to do so (Dunlosky & Hertzog, 2001;

see Kausler, 1994, for a review).

In contrast to the above LTM related production deficits, no such age-differences in the proportion of self-reported elaboration has been observed in WM span tasks (Bailey et al., 2009). Both older and young adults reported using mental imagery (OA: 11%, YA: 14%), and sentences (OA: 12%, YA: 8%) comparatively frequently across the trials of two WM span tasks. Although the use of these strategies again correlated positively with WM span performance, it did not account for age-related variance in WM span performance (Bailey et al., 2009). Experimental investigations of the effect of elaboration on WM performance showed no age-difference in the null-effect of elaboration: The instruction to form a vivid mental image of parts of the memoranda after list presentation did not benefit WM, in neither young nor older adults (Bartsch, Loaiza, Jäncke, et al., 2019; Bartsch & Oberauer, 2019).

Taken together, LTM deficits in older adults might at least in part arise from a deficiency in generating elaborations at encoding. Therefore, the second goal of the present study was to investigate why older adults show so little benefit of elaboration, and specifically, to test the generation-deficit hypothesis. If this hypothesis was correct, then providing older adults with sentences that enrich the to-be-remembered words should help them overcome the generation deficit, so that their LTM benefits from elaboration as much as young adults do. Furthermore, extending the first goal of the experiment, we aimed to investigate whether elaboration of information in WM, when facilitated by providing enriching information, improves WM performance in older adults as well.

The present study

Here, we bring together three approaches to studying the effects of elaboration on WM as well as LTM functioning: (a) strategy reports, (b) the benefits of free time, and (c) experimentally instructing elaboration, to answer three questions: (1) What strategies people spontaneously use during free time, (2) whether young and older adults benefit from assisted elaboration through sentences in a test of WM, and (3) whether the LTM deficit in older adults arises in part from a deficit in generating the enrichment of the memory material necessary for effective elaboration.

We tested WM through immediate serial recall of a list of nouns. To assist elaboration, in the *sentence* conditions the nouns were interleaved by filler words embedding them into meaningful sentences. In one of the two sentence conditions, we additionally instructed participants to form a mental image of the meaning of the sentence. These two conditions were compared to two baseline conditions without interleaved words, which differed in the amount of free time in between the nouns. In the *short baseline* condition, the time for presenting the interleaved words was cut out, whereas in the *long baseline* condition these time intervals were retained as free time. To gauge the strategies people spontaneously used during free time, and to check that they followed our instructions in the sentence conditions, in Experiment 1we asked them after each trial which strategy they had used on that trial.

We expected that the longer free time should lead to better serial recall compared to the short baseline (i.e. a free time benefit). If spontaneous elaboration underlies this free-time benefit for serial-order WM, we predict that participants report more elaboration in the long than the short baseline in Experiment 1. As elaboration has never been reported as a spontaneous strategy by a majority of participants, there is ample room for increasing its prevalence further through assisted elaboration and instruction in the sentence conditions. We therefore predict that participants report elaboration much more frequently in the sentence conditions compared even to the long baseline. If this is the case, we can answer our second research question through the following prediction: If elaboration is beneficial for WM, then immediate serial recall in the sentence condition should be better than in the *long baseline* condition.

We further included in all three experiments a delayed memory test of the nouns that had been presented as memoranda for the immediate serial-recall trials as a manipulation check: The extent to which our manipulation of elaboration increases the degree to which people engage in elaboration (relative to the other conditions) should be reflected in an increase of LTM, at least in young adults. Furthermore, if spontaneous elaboration underlies the free-time benefit in WM, the effect of our experimental manipulations on delayed recall should mirror those on the WM test. Specifically, if – as we expect – providing sentences and imagery instruction leads to more elaboration than people spontaneously engage in in the *long-baseline* condition, then performance should exceed that in the long baseline condition for both the WM and the LTM tests. Conversely, if spontaneous elaboration was not responsible for the free-time benefit in WM, then assisting elaboration by providing sentences should improve LTM but not WM relative to the *long baseline.*

The delayed memory test further served to answer our third question, whether the LTM deficit in older adults arises in part from a deficit in generating the enrichment of the memory material necessary for effective elaboration: To this end, in Experiments 2 and 3 we compared a group of young to a group of older adults. On the generation-deficit hypothesis we expect that, when older adults are provided with sentences that enrich the memory material for them, their LTM should benefit from elaboration to a comparable degree as that of young adults.

Method

Participants

For Experiment 1 we recruited 24 Students (15 female) from the University of Zurich. For Experiment 2 we recruited 24 students (13 female) from the University of Zurich and 24 healthy older adults (14 female) from the Zurich community as participants. Participants in Experiment 3 were 21 students (13 female) from the University of Zurich and 20 healthy older adults (10 female) from the Zurich community.

In all three studies, participants were compensated with either 15 Swiss Francs (about 15 USD) or partial course credit for the one-hour experiment. The studies were carried out in agreement with the rules of the Ethics Committee of the Faculty of Arts and Sciences of the University of Zurich.

Cognitive functioning was screened with the MMSE (Mini-Mental Status Examination; Folstein, Folstein, & McHugh, 1975), indicating age-typical cognitive abilities in our sample of older adults (Experiment 2: M = 29, SD = 1.14, range = 26–30; Experiment 3: M = 28.86, SD = 1.38, range = 25–30). Table 1 shows the descriptive statistics and

Table 1
Sample Description (means (and standard deviations)) of Experiment 1, 2 and 3.

Experiment	Age Group	Age	vocabulary	digit-symbol task	calibrated pres. Time in ms
1	Younger	24.92 (3.71)	29.75 (4.19)	67.08 (10.69)	500.00 (183.74)
2	Younger	22.46 (3.05)	29.81 (2.24)	67.48 (8.27)	481.25 (145.66)
	Older	70.9 (4.3)	32.38 (1.63)	46.5 (8.19)	604.49 (113.35)
	$PD_{age\text{-}effect}$	–	100% < 0 < 0%	0% < 0 < 100%	99.8% < 0 < 0.2%
3	Younger	24.67 (3.15)	27.64 (4.86)	66.71 (8.89)	462.5 (134.23)
	Older	70.76 (3.65)	33.81 (7.57)	49.05 (9.49)	656.83 (115.29)
	$PD_{age\text{-}effect}$	–	99.3% < 0 < 0.7%	0% < 0 < 100%	100% < 0 < 0%

Note. PD is the posterior density of the age effects. Zero represents the point of no age differences, and the percentages indicate how much of the estimated effect's posterior distribution lies below and above 0. Values below 0 reflect higher values of older adults whereas positive values indicate higher values of younger adults.

posterior distributions of the age effects of our sample. The evidence indicates slower processing speed in the older compared with the young adults as measured by the digit-symbol test (Petermann & Wechsler, 2012). The older adults showed better performance than the young adults in a computerized vocabulary test (Mehrfachwahl-Wortschatz Test Version B, Lehrl, 2005), consisting of 37 items in which participants are supposed to find an existing word among four similarly sounding non-words. The MWT-B is a marker test for crystallized intelligence. Hence, our sample of young and old adults showed typical age differences in processing speed and measures of crystallized intelligence (Li et al., 2004).

Materials and procedure

In the three experiments presented here, we asked participants to remember short lists of nouns in serial order. The stimuli were drawn from a pool of 450 German concrete nouns. The nouns were between three and nine letters long and had a mean normalized lemma frequency of 23.41/million (drawn from the dlexdb.de lexical database).

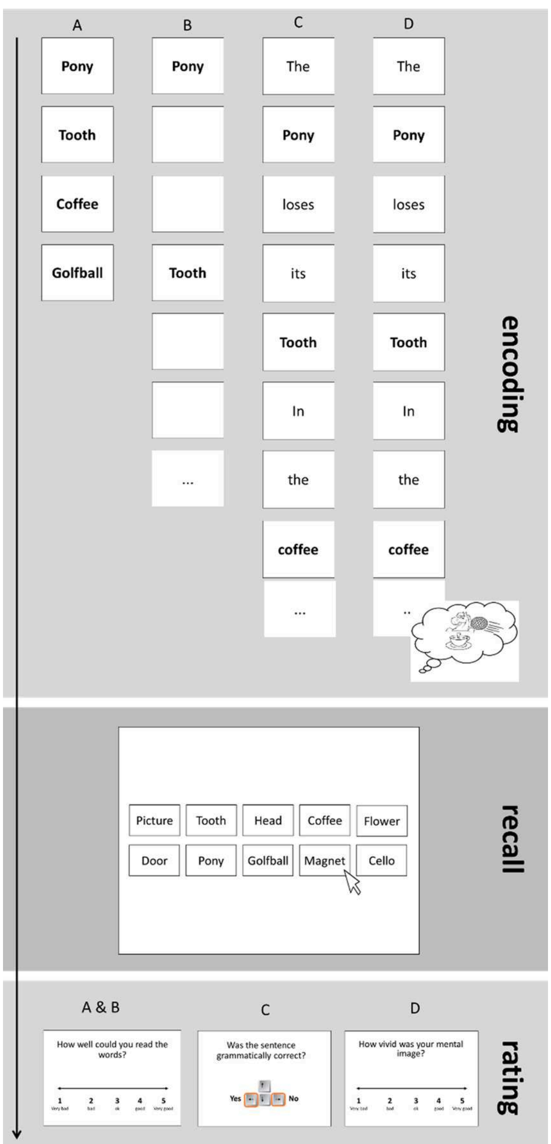

Fig. 1. Illustration of the working memory paradigm. Subjects were shown a list of words sequentially according to the four experimental conditions: A) the short baseline, B) the long baseline C) the sentence grammar condition and D) the sentence imagery condition.

The sequence of an experimental trial is illustrated in Fig. 1. The presentation rate of each word was self-adjusted by each participant at the beginning of the experiment: Example sentences were presented word-by-word centrally on the screen, and subjects were asked to adjust the presentation time, so they were comfortably able to read the sentences. In the main experiment, participants were informed prior to each trial about the following experimental condition. After sequential presentation of the to-be-remembered words, a WM test followed immediately: An array of words was displayed, and participants were to reconstruct the memory list by clicking on the list words in their serial order. The response options in the array consisted of all of the words of that trial's memory list, and the same number of new items. The position of the options on the screen was random, and participants used the mouse to select among them at their own pace. The immediate memory test was chosen, because it is a common test for WM that enables strong control over the response set from which the responses are chosen. In this way, we control and hold constant the demand on item memory (choosing one of the six list items out of 12 options) and on relational memory (choosing the correct item out of six list items for the current list position) (Oberauer & Lewandowsky, 2019).

There were four encoding conditions. In the *short baseline* condition, the to-be-remembered words were presented individually in the center of the screen for the amount of time taken from the adjustment period (e.g. 500 ms), interleaved by a short ISI of 100 ms. In the *long baseline* each word was followed by the presentation of a blank screen for two times the word presentation time (e.g. 1000 ms), equivalent to the time to present two filler words. For the *sentence grammar* and *sentence imagery* conditions, the to-be-remembered words were presented within a sentence, each word followed by, on average, two filler words. To make the memoranda (i.e., the nouns) very distinct to the participant, the nouns were presented in bold, fillers were never nouns, and the fillers were never included among the response options. Five independent subjects were invited to the lab to create the sentences, and the first author selected 75 sentences to be used in the experiments, based on their meaningfulness and in accordance with the desired 2:1 ratio of filler words to target nouns. In both sentence conditions, all the words were presented centrally on the screen at the pace adjusted by each individual in the beginning of the experiment. In the *sentence grammar* condition, the subjects were asked to judge whether the sentence was grammatically correct, which they were in 50% of the cases. In the *sentence imagery* condition, the subjects were asked to form a vivid mental image of the scene described in the sentence. Following the memory test, the subjects were to rate the vividness of the created mental image (*sentence imagery* condition), the grammatical correctness of the sentence (*sentence grammar* condition), or how well they were able to read the words (*short & long baseline* condition).

There were four trials of the WM task per block, one of each condition. The experiment comprised eight blocks. An unrelated distracter task followed each block, in which the participants had to indicate the correctness of visually presented math equations (e.g. $9 \times 8 = 72$) for 2 min. After that followed a typed delayed free recall memory test, wherein the participants were asked to recall as many memory items from the previous block as possible. At each of these eight delayed-recall tests subjects were to recall all the items of the past four trials of the WM task, which – depending on the set size of the respective experiments – meant to recall 20 (Exp. 1 & 2), 16 (OA in Exp 3) or 24 (YA in Exp 3) words in total. This test served to assess the effect of each experimental condition on episodic LTM. The participants were made aware of the delayed memory test before the start of the experiment. We chose free recall for the LTM test because it is a common procedure for assessing the effect of elaboration on episodic LTM.

With Experiment 1 we aimed to assess what processes they reportedly engaged in (more) during the long baseline compared to the short baseline condition at memory set size of five nouns, and to gauge whether they followed our instructions in the sentence conditions.

Participants indicated per mouse-click after each trial whether they engaged in passive reading, rote repetition, use of sentences, imagery, the combined use of sentences and imagery, or meaningful grouping (Strategies were adapted from: Bailey et al., 2009; Bailey, Dunlosky, & Kane, 2008; Dunlosky & Kane, 2007).

In Experiment 2, we included both a group of young and older adults to perform the immediate serial recall task with a memory set size of five nouns. In this way we could compare performance of the two age groups at the same level of load on WM. The same nominal memory load, however, poses a higher demand on WM for old adults who have, on average, a lower WM capacity than the young. This could impair their ability to form a robust, accessible trace of the memoranda in episodic LTM, and in particular it could compromise their ability to form an integrated representation of the meaning of the given sentence, thereby undercutting any potential benefit of elaboration we aimed to induce. Therefore, we ran Experiment 3, which was identical to Experiment 2, but varied memory set size between the age-groups (6 nouns for young vs. 4 nouns for older adults). The reduced memory load for old adults served to compensate for their reduced WM capacity.[1]

Data analysis

We analyzed the data using Bayesian generalized linear mixed models (BGLMM) implemented in the R package *brms* (Bürkner, 2017, 2018).

For the data of the strategy report in Experiment 1 we ran separate models for each strategy category, with condition as independent variable, and the proportion of trials on which a person chose that strategy as dependent variable. Here, we assumed a Gaussian data distribution predicted by a linear model through an identity link function. The regression coefficients were given weakly informative student-t priors with three degrees of freedom and a scaling parameter of 10.

For the data of the WM task the dependent variable was the correctness (0 or 1) of each of the responses in each trial (serial recall of first to last item) of each condition per participant. Correct responses were defined as choosing the target item from the alternatives (i.e., all other list items and new items). Therefore, we assumed a Bernoulli data distribution predicted by a linear model through a logit link function (i.e., a repeated-measures logistic regression). The regression coefficients were given weakly informative Cauchy priors with a mean of 0 and the standard deviation of 5.

For the data of the LTM task the dependent variable was the proportion of correctly recalled items in each condition per block and participant. Correct responses were defined as recalling one of the nouns that was presented in the previous block of WM trials. Here, we assumed a Gaussian data distribution predicted by a linear model through an identity link function. The regression coefficients were given weakly informative student-t priors with three degrees of freedom and a scaling parameter of 10.

For all analyses, following the recommendation of Barr and colleagues (Barr, Levy, Scheepers, & Tily, 2013 see also Schielzeth & Forstmeier, 2009) we implemented the maximal random-effects structure justified by the design. Here, this included a by-participant random intercept and a random slope for condition. In addition, we estimated the correlation among the random-effects parameters. We used completely non-informative priors for the correlation matrices, so-called LKJ priors with shape parameter 1.

Bayesian procedures provide posterior probability distributions of the model parameters (i.e., the regression weights) that express uncertainty about the estimated parameters. The highest density regions (HDRs) of these posteriors can be used for statistical inference. A 95% HDR represents the range in which the true value of a parameter lies with probability 0.95, given model and data (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). If zero lies outside the Bayesian HDR there is strong evidence for the existence of the corresponding effect; although the strength of evidence varies continuously, for simplicity we will describe effects as "credible" if their HDRs exclude zero.

We used an MCMC algorithm (implemented in Stan; Carpenter et al., 2017) that estimated the posteriors by sampling parameter values proportional to the product of prior and likelihood. These samples are generated through 4 independent Markov chains, with 500 warmup samples each, followed by 1000 samples drawn from the posterior distribution which were retained for analysis. Following Gelman et al. (2013), we confirmed that the 4 chains converged to the same posterior distribution by verifying that the $\widehat{R}$ statistic – reflecting the ratio of between-chain variance to within-chain variance – was <1.06 for all parameters, and we visually inspected the chains for convergence.

Because we use Bayesian statistics, we did not use power considerations for deciding on our sample size – the concept of power is not defined in Bayesian statistics. What best corresponds to it is the precision of the posterior estimates of standardized effect sizes. We based our sample sizes on our previous studies on this topic (Bartsch, Loaiza, Jäncke, et al., 2019; Bartsch & Oberauer, 2019), which yielded sufficiently precise posteriors to permit confident inferences.

Results

In the following, we report the results of all experiments jointly in light of our three research questions: (1) Which strategies do subjects report to engage in naturally? (2) How does elaboration affect WM and LTM performance? (3) And does providing enriched information overcome older adults' elaboration deficit? First, we address which strategies were reported in Experiment 1, with a focus on which processes subjects report to engage in (more) during the long baseline compared to the short baseline condition. We further ask whether the use of elaborative strategies resulted in subsequent memory effects. We then turn to the experimental effects of Experiments 1–3, and ask how the experimental manipulation of elaboration affected WM. Those effects are further analyzed in interaction with the strategy reports as well as in interaction with age.

The same approach is taken for the data of the LTM task, with the additional question whether providing enriched information overcome the older adults LTM deficit. All data and analysis scripts can be assessed on the Open Science Framework (https://osf.io/4n9y3).

Reading times

The mean self-chosen reading speed of Experiment 1 was 500 ms (SD = 183.74 ms) per word for the young adults. The mean self-chosen reading speed of young adults in Experiment 2 and 3 was 481 ms (SD = 145) and 462.5 ms (SD = 135.23) per word. For the older adults, the mean self-chosen reading speed was 604 ms (SD = 113) and 656.84 ms (SD = 115.29) for Experiment 2 and 3.

Which strategies are reported?

The mean proportion of the reported strategies of Experiment 1 is shown in Fig. 2. The posterior effect estimates can be seen in Table 2. As a first step we were interested in the compliance of the subjects to read the sentences (sentence grammar condition) and additionally forming a mental image in the sentence imagery condition. The self-report data show that subjects indeed increased the use of the sentences in both conditions compared to the baselines (baselines vs. sentence conditions $\Delta = 0.25$, 95% HDR = [0.19, 0.31]), and additionally increased the proportion of trials in which they were forming a mental image in the sentence imagery compared to the sentence grammar condition

[1] Experiment 1 was carried out after Experiments 2 and 3. Because of the Covid-19 pandemic we could not invite older adults into the lab, and therefore did not include an older sample in this experiment.

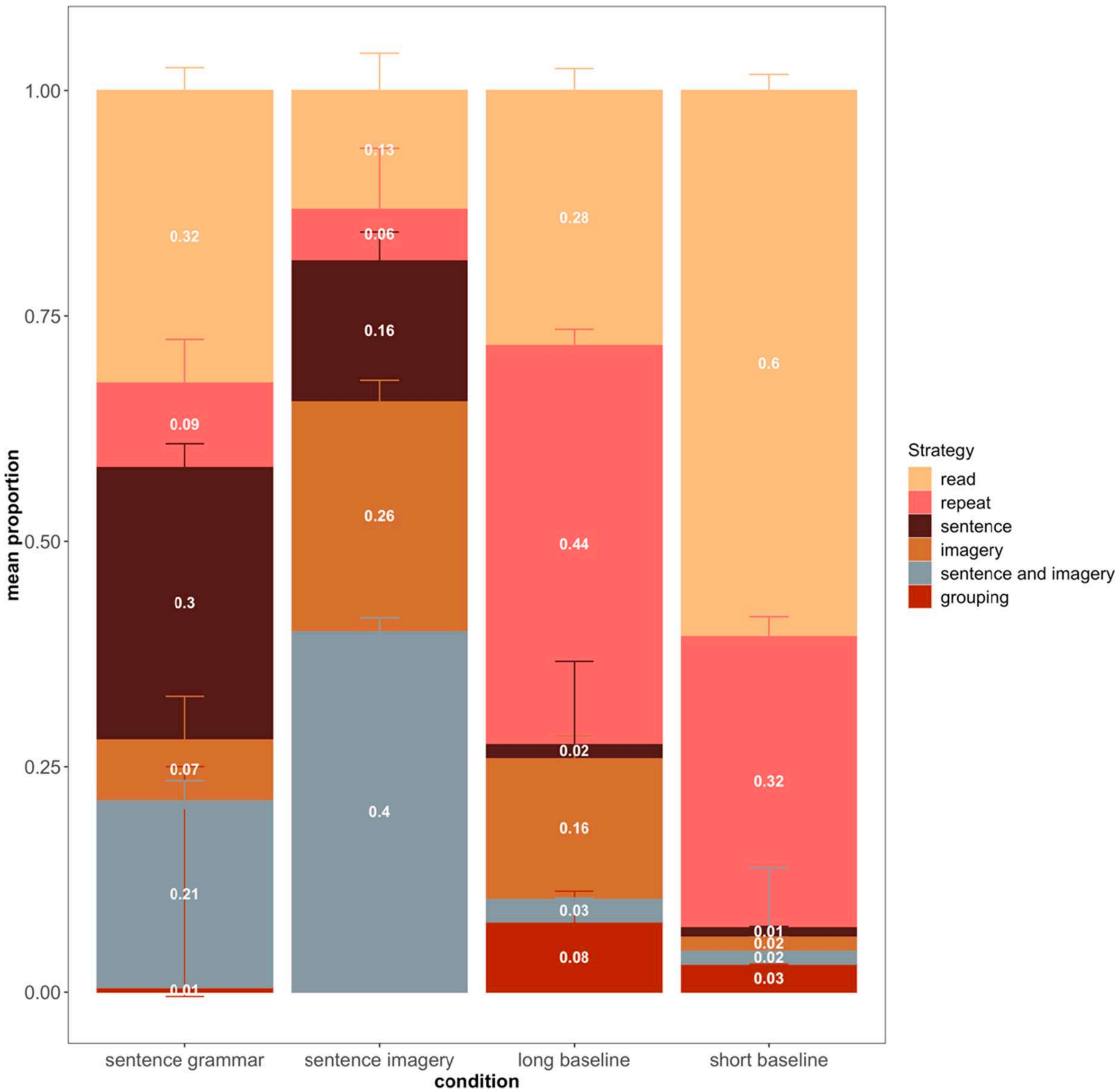

Fig. 2. Mean proportions of reported strategies in each of the four conditions of Experiment 1. The error bars reflect standard error of the mean.

(sentence imagery vs. sentence grammar condition $\Delta = 0.19$, 95% HDR $= [0.06, 0.32]$).

Furthermore, we were interested in the strategies subjects reportedly engaged in *more* during the long baseline compared to the short baseline condition. The BGLMM of reported strategies across the two baseline conditions revealed that the mean proportion of reading decreased (short vs. long baseline condition $\Delta = -0.32$, 95% HDR $= [-0.48, -0.18]$), and the proportion of trials with reported mental imagery *increased* with longer free time (short vs. long baseline condition $\Delta = 0.14$, 95% HDR $= [0.05, 0.26]$).

In conclusion, subjects followed our instructions and increased the use of elaborative strategies in both sentence conditions, and especially increased the use the of mental imagery in the sentence imagery condition. In the long baseline, subjects reportedly engaged more in rote rehearsal than in the short baseline, but also increased their use of mental imagery, which strengthens our motivation to assess whether the latter increase drives the free time benefit.

Effect of elaborative vs. non-elaborative strategies on performance

In order to assess the effectiveness of the engagement in elaborative strategies we compared the subsequent memory performance of trials in which these were reported to trials in which non-elaborative strategies were reported. Following (Bailey et al., 2009; Dunlosky & Kane, 2007), we considered use of sentences, imagery, combined use of sentences and imagery, and meaningful grouping to be elaborative, whereas passive reading and rote repetition were considered to be non-elaborative. Using these two categories, we analyzed immediate and delayed recall performance as a function of experimental condition, and of elaborative vs. non-elaborative reported strategies.

Fig. 3A shows the mean serial-recall performance and their corresponding 95% HDR's in the working-memory task of Experiment 1. The posterior effect estimates of the model parameters (i.e., the regression weights) are presented in Table 4. The BGLMM revealed evidence for a main effect of type of strategy (elaborative vs. non-elaborative), with higher performance when subjects reportedly had engaged in elaborative strategies in the WM task across conditions (elaborative vs. non-elaborative $\Delta = 0.61$, 95% HDR $= [0.16, 1.09]$). More specifically, the BGLMM revealed better immediate serial memory performance for when subjects reportedly had engaged in elaborative strategies in trials of the *sentence imagery* as well as in the *long and short baseline* condition ($\Delta = -1.36$, 95% HDR $= [-2.14, -0.59]$; $\Delta = -0.9$, 95% HDR $= [-1.82, -0.04]$ and $\Delta = -2.37$, 95% HDR $= [-4.39, -0.86]$, respectively).

Table 2
The posterior effect estimates of the pairwise contrasts and their 95% HDRs of the generalized linear mixed model for the strategy report data of Experiment 1.

contrast	strategy	Mode of parameter on logit scale	95% HDR
long baseline vs. short baseline	**read**	**−0.32**	**[−0.48, −0.18]**
	repeat	**0.12**	**[0.01, 0.24]**
	mental imagery	**0.14**	**[0.05, 0.26]**
	sentence use	0.01	[−0.01, 0.02]
	sentence & mental imagery	0.01	[−0.02, 0.04]
sentence imagery vs. long baseline	read	−0.15	[−0.29, 0.00]
	repeat	**−0.38**	**[−0.50, −0.27]**
	mental imagery	0.10	[−0.08, 0.26]
	sentence use	**0.14**	**[0.05, 0.24]**
	sentence & mental imagery	**0.38**	**[0.24, 0.49]**
sentence grammar vs. long baseline	read	0.04	[−0.09, 0.17]
	repeat	**−0.35**	**[−0.48, −0.21]**
	mental imagery	−0.09	[−0.19, 0.02]
	sentence use	**0.29**	**[0.16, 0.42]**
	sentence & mental imagery	**0.18**	**[0.07, 0.30]**
sentence grammar vs. sentence imagery	**read**	**0.19**	**[0.07, 0.31]**
	repeat	0.04	[−0.03, 0.10]
	mental imagery	**−0.19**	**[−0.31, −0.07]**
	sentence use	**0.15**	**[0.02, 0.27]**
	sentence & mental imagery	**−0.19**	**[−0.32, −0.06]**
sentence grammar vs. short baseline	**read**	**−0.28**	**[−0.46, −0.11]**
	repeat	**−0.23**	**[−0.33, −0.11]**
	mental imagery	0.05	[0.00, 0.10]
	sentence use	**0.29**	**[0.17, 0.42]**
	sentence & mental imagery	**0.19**	**[0.08, 0.31]**
sentence imagery vs. short baseline	**read**	**−0.47**	**[−0.64, −0.30]**
	repeat	**−0.26**	**[−0.38, −0.15]**
	mental imagery	**0.24**	**[0.11, 0.37]**
	sentence use	**0.15**	**[0.06, 0.24]**
	sentence & mental imagery	**0.39**	**[0.26, 0.51]**

Note. Credible differences, defined as HDRs excluding zero, are printed in bold. Negative parameter estimates imply that a strategy was reported less often in the first condition of a contrast; positive estimates imply that the strategy was reported more often in the first condition of the contrast.

Self-reported elaboration was also associated with better episodic LTM, but the evidence for that main effect was weaker and not credible by our criterion: (Δ = 0.08, 95% HDR = [−0.01, 0.16]) (see Table 5). Yet, as shown in Table 5, the posterior effect estimates of the BGLMM revealed better delayed memory performance for when subjects reportedly had engaged in elaborative strategies in trials of the *sentence imagery* as well as in the *sentence grammar* condition (Δ = −0.35, 95% HDR = [−0.46, −0.25] and Δ = −0.11, 95% HDR = [−0.21, −0.01], respectively). This was not the case for the long baseline condition (Δ = 0.0309, 95% HDR = [−0.07, 0.14]), as can be seen also in Fig. 3B.

Taken together, our analysis showed that words that were initially reported to have been processed with an elaborative strategy were recalled with a higher probability in the immediate and, to some extent, the delayed test.

Working memory

How does the experimental manipulation of elaboration affect WM of young adults?

Fig. 3A, Fig. 4A and Fig. 5A show the mean serial-recall performance and their corresponding 95% HDR's in the working-memory task of Experiment 1, 2 and 3, respectively. The posterior effect estimates of the model parameters (i.e., the regression weights) are presented in Table 3, Table 6 and Table 8.[2]

Our first question was whether our manipulation of free time between the presentation of items in a memory list replicated the usual effect on immediate serial recall (WM task). There was a credible difference between the short and long baseline for young adults in all three Experiments, implying that participants had better memory for items interleaved with a longer free-time interval than for items without this free time (see Table 3, Table 6 and Table 8).

Next, we were interested in how the sentence conditions affected performance compared to the two baselines. The BGLMM revealed that performance of young adults in the sentence-imagery condition approximated that in the long baseline. Yet, the comparison to the short baseline revealed that performance in the sentence imagery condition was not credibly better either. Immediate serial recall in the sentence-grammar condition was poorer than the long baseline in Experiments 2 and 3, suggesting that the process of evaluating the sentence's grammaticality was not an effective form of elaboration for WM. Taking a closer look at the experimental effects depending on the self-reported strategy in Experiment 1, we found that the difference between short and long free time was credible only with non-elaborative strategies. Furthermore, with elaborative strategies, sentence imagery resulted in performance equal to the long baseline, but with non-elaborative strategies, it led to worse performance (see Fig. 3A and Table 4).

Does assisted elaboration affect older adults' WM similar to young adults'?

There was a main effect of age in Experiment 2, reflected in the effects' HDR excluding zero, with younger adults outperforming the older adults in the WM task across conditions (older vs. young Δ = −0.88, 95% HDR = [−1.33, −0.44]). As the older adults were presented with only four words in Experiment 3, there they outperformed the younger adults in the WM task across conditions (older vs. young Δ = 1.13, 95% HDR = [0.64, 1.64]).

Equivalently to the young adults, older adults benefited from free time and showed credibly better performance in the long compared to the short baseline in Experiment 2 (see Table 6). Yet, in Experiment 3, old adults' performance might have been too close to ceiling to allow a sizeable free-time benefit (see Table 8). The immediate serial recall in the sentence-grammar condition was poorer than the long baseline for both age groups, suggesting that the process of evaluating the sentence's grammaticality was not an effective form of elaboration for WM also in older adults. For the older adults, sentence imagery was even less helpful than for the young in both Experiment 2 and 3: Their performance in the sentence imagery condition was poorer than in the long baseline, and even somewhat worse than the short baseline.

[2] There was credible evidence for all of the random effects. The interested reader can find the respective posterior effect estimates in Supplementary Table 2 and 3.

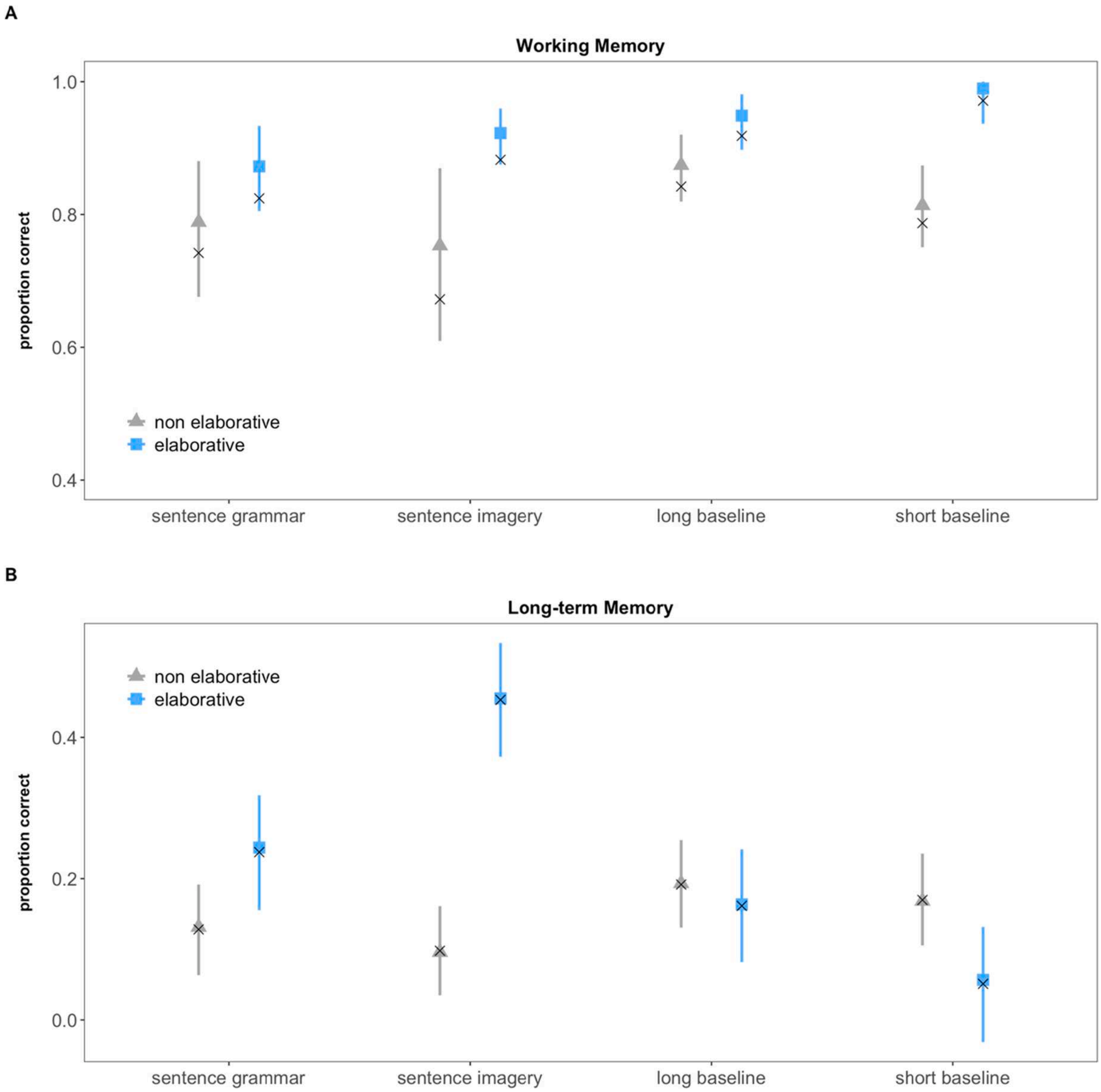

Fig. 3. Proportion correct in the (A) working-memory task and (B) long-term memory task in Experiment 1. The blue (elaborative strategies) and grey (non-elaborative strategies) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMMs. The crosses represent the observed proportions. Their overlap indicates that the models adequately describe the data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Long-term memory

How does the experimental manipulation of elaboration affect LTM of young adults?

Fig. 3B, Fig. 4B and Fig. 5B show the mean free recall performance and their corresponding 95% HDR's in the delayed memory task of Experiment 1, 2 and 3. The posterior effect estimates are presented in Table 3, Table 7 and Table 9.[3] The first comparison of interest concerned assessing to what extent our manipulation of elaboration benefited delayed memory performance. We therefore compared performance in the sentence imagery condition to both baselines. As seen in Table 3, Table 7 and Table 9, for young adults the sentence imagery condition yielded better performance than both baseline conditions in all three experiments, demonstrating that assisted elaboration effectively boosted episodic LTM, even more so than the long baseline. Moreover, the younger adults showed higher delayed recall performance in the sentence imagery condition than the sentence grammar condition, suggesting that forming an image of the sentence's meaning was a more effective elaboration process than evaluating the sentence's grammaticality.

Taking a closer look at the experimental effects on delayed memory depending on the self-reported strategy in Experiment 1, we found that, in contrast to immediate memory, the beneficial effect of long compared to short free time on LTM was credible only with *elaborative* strategies. Furthermore, with elaborative strategies, sentence imagery resulted in better performance than the long baseline, but with non-elaborative strategies, the sentence-imagery condition led to even worse LTM than the long baseline (see Fig. 3B and Table 5).

In conclusion, we showed a large benefit of our elaboration manipulation on LTM and thereby ensured, that assisting elaboration through sentences and instructed imagery increased the amount or effectiveness

[3] There was credible evidence for parts of the random effects. The interested reader can find the respective posterior effect estimates in Supplementary Table 3.

Table 3
The posterior effect estimates of the pairwise contrasts of the main effect of condition and their 95% HDRs of the generalized linear mixed model for the immediate serial and delayed memory data of Experiment 1.

contrast	memory task	Mode of parameter on identity scale	95% HDR
long baseline vs. short baseline	**immediate**	**0.45**	**[0.01, 0.89]**
	delayed	0.07	[−0.01, 0.15]
sentence imagery vs. long baseline	immediate	0	[−0.57, 0.58]
	delayed	**0.1**	**[0.02, 0.18]**
sentence grammar vs. long baseline	immediate	−0.46	[−0.99, 0.05]
	delayed	0.01	[−0.07, 0.08]
sentence grammar vs. sentence imagery	immediate	−0.46	[−1.05, 0.09]
	delayed	**−0.09**	**[−0.17, −0.01]**
sentence grammar vs. short baseline	immediate	0	[−0.48, 0.44]
	delayed	0.07	[−0.01, 0.15]
sentence imagery vs. short baseline	immediate	0.45	[−0.12, 1.04]
	delayed	**0.17**	**[0.08, 0.25]**

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 4
The posterior effect estimates of the pairwise contrasts and their 95% HDRs of the generalized linear mixed model for the immediate serial memory data of Experiment 1.

contrast	strategy	Mode of parameter on identity scale	95% HDR
long baseline vs. short baseline	elaborative	−0.96	[−2.97, 0.49]
	non-elaborative	**0.43**	**[0.01, 0.88]**
sentence imagery vs. long baseline	elaborative	−0.36	[−1.24, 0.45]
	non-elaborative	**−0.85**	**[−1.52, −0.15]**
sentence grammar vs. long baseline	**elaborative**	**−0.88**	**[−1.73, −0.09]**
	non-elaborative	**−0.63**	**[−1.25, −0.02]**
sentence grammar vs. sentence imagery	elaborative	−0.5	[−1.13, 0.1]
	non-elaborative	0.22	[−0.49, 0.89]
sentence grammar vs. short baseline	**elaborative**	**−1.91**	**[−3.86, −0.46]**
	non-elaborative	−0.18	[−0.76, 0.38]
sentence imagery vs. short baseline	elaborative	−1.45	[−3.36, 0.1]
	non-elaborative	−0.39	[−1.05, 0.27]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

of elaboration beyond that achieved in the long baseline.

Does providing enriched information overcome older adults' elaboration deficit?

For our third question, we looked at the effects of age on the delayed memory data. As seen in Fig. 4B and supported by evidence for a main effect of age (older vs. young: $\Delta = -0.15$, 95% HDR = $[-0.23, -0.06]$) the older adults remembered less words than the younger adults in the delayed test of Experiment 2. Their performance was equivalent across all conditions, including assisted elaboration through sentences. Fig. 6 depicts the interaction effect of condition by age group, showing evidence that the beneficial effect of sentence imagery – in comparison to the two baselines and to the sentence-grammar condition – was larger for young than for old adults.

Table 5
The posterior effect estimates of the pairwise contrasts and their 95% HDRs of the generalized linear mixed model for the delayed memory data of Experiment 1.

contrast	strategy	Mode of parameter on proportion-correct scale	95% HDR
long baseline vs. short baseline	**elaborative**	**0.11**	**[0.04, 0.18]**
	non-elaborative	0.02	[−0.05, 0.1]
sentence imagery vs. long baseline	**elaborative**	**0.3**	**[0.22, 0.37]**
	non-elaborative	**−0.1**	**[−0.17, −0.02]**
sentence grammar vs. long baseline	elaborative	0.08	[0, 0.15]
	non-elaborative	−0.06	[−0.14, 0.01]
sentence grammar vs. sentence imagery	**elaborative**	**−0.22**	**[−0.29, −0.14]**
	non-elaborative	0.03	[−0.05, 0.11]
sentence grammar vs. short baseline	**elaborative**	**0.19**	**[0.11, 0.26]**
	non-elaborative	−0.04	[−0.12, 0.03]
sentence imagery vs. short baseline	**elaborative**	**0.4**	**[0.32, 0.48]**
	non-elaborative	−0.07	[−0.15, 0]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

As the older adults' WM performance in Experiment 2 was already lower than that of the younger adults, we made the task easier for them and decreased memory load in Experiment 3. Now that overall performance in the WM task was higher for the older than the young adults, thereby more than compensating for the age deficits at the WM stage, old adults should have been able to fully process the sentences provided, thereby making optimal use of the assisted elaboration. Fig. 5B shows that, whereas young adults in Experiment 3 again benefited from elaboration, older adults did not.

As supported by evidence for a main effect of age (older vs. young: $\Delta = -0.10$, 95% HDR = $[-0.17, -0.03]$) the older adults remembered a lower proportion of words than the younger adults, and their performance was equivalent across all conditions, including assisted elaboration through sentences. Fig. 7 depicts the interaction effect of condition by age group, showing evidence that, relative to both baselines, sentence imagery improved episodic memory more for young than for old adults.

Discussion

The first goal of the present study was to examine to what extent young and older adults can benefit from assisted elaboration through sentences in a test of WM as well as a test of LTM, and to study elaboration as a potential cause for the free-time benefit in WM. We assumed, that if free time is spontaneously used to engage in elaboration by some participants sometimes, and if elaboration is beneficial for both WM and episodic LTM, then assisted elaboration should improve memory above the long-baseline level in both WM and LTM. This should be the case in young adults - who we assume are good at elaborating - and even more for old adults who have been shown to be deficient in generating elaborations themselves, and therefore should gain more from external assistance to elaboration.

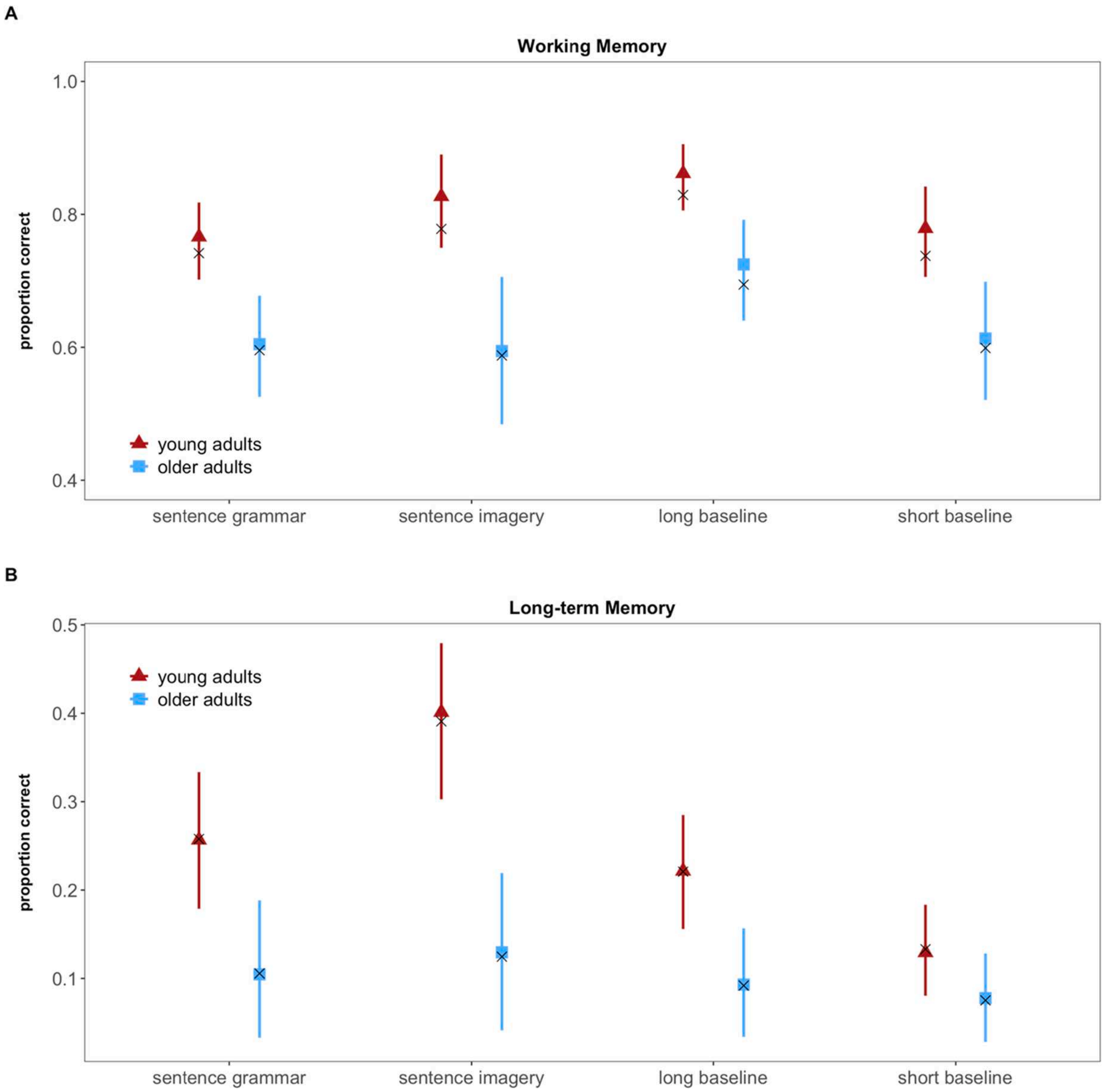

Fig. 4. Proportion correct in the (A) working-memory task and (B) long-term memory task in Experiment 2. The red (young adults) and blue (older adults) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMMs. The crosses represent the observed proportions. Their overlap indicates that the models adequately describe the data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Elaboration and the free time benefit on working memory

Our results confirm that young adults benefit from free time interleaving the to-be-remembered items in WM, and so do older adults when performance is not close to ceiling. One potential explanation for this effect is that people use free time to elaborate the memoranda, and that helps immediate recall.

We know from self-report studies that only one fourth of the subjects indicate to spontaneously elaborate in WM tasks (e.g. Bailey et al., 2011; Dunlosky & Kane, 2007). Experiment 1 confirmed this modest proportion of spontaneous elaboration: The increased free time in the long baseline was used in around 16% of the trials for mental imagery, and an additional 5% of trials were accompanied by sentence generation alone or in combination with imagery.

We therefore expected that instructing *all* participants to form a mental image of the memoranda should have led to more consistent elaboration than the long baseline. Additionally, the fact that we also assisted elaboration by providing meaningful sentences should have boosted the effectiveness of elaboration. The increased frequency of reporting the use of elaborative strategies in these conditions in Experiment 1, as well as the effects of this manipulation on LTM in the young-adult group, corroborated that assumption: Both sentence conditions resulted in a substantial benefit for LTM compared to the short baseline, and in case of the sentence imagery condition also compared to the long baseline. This implies that, compared to the long baseline condition, in the sentence imagery condition more people engaged in elaboration, or they did so more effectively.

These findings demonstrate that we managed to increase the amount or effectiveness of elaboration, at least in young adults. Then the critical question was, how this affected participants' WM. We found that the sentence imagery condition never surpassed the long baseline in young adults (and in fact always ended up a bit short). The older adults, instead of being able to compensate their elaboration deficit, performed more poorly with assisted elaboration than in the long baseline. These two findings question the idea that the free-time benefit on WM is to a large extent due to elaboration. This conclusion is further bolstered by the observation from Experiment 1 that longer free time improved WM only for the subset of trials for which participants reported non-elaborative

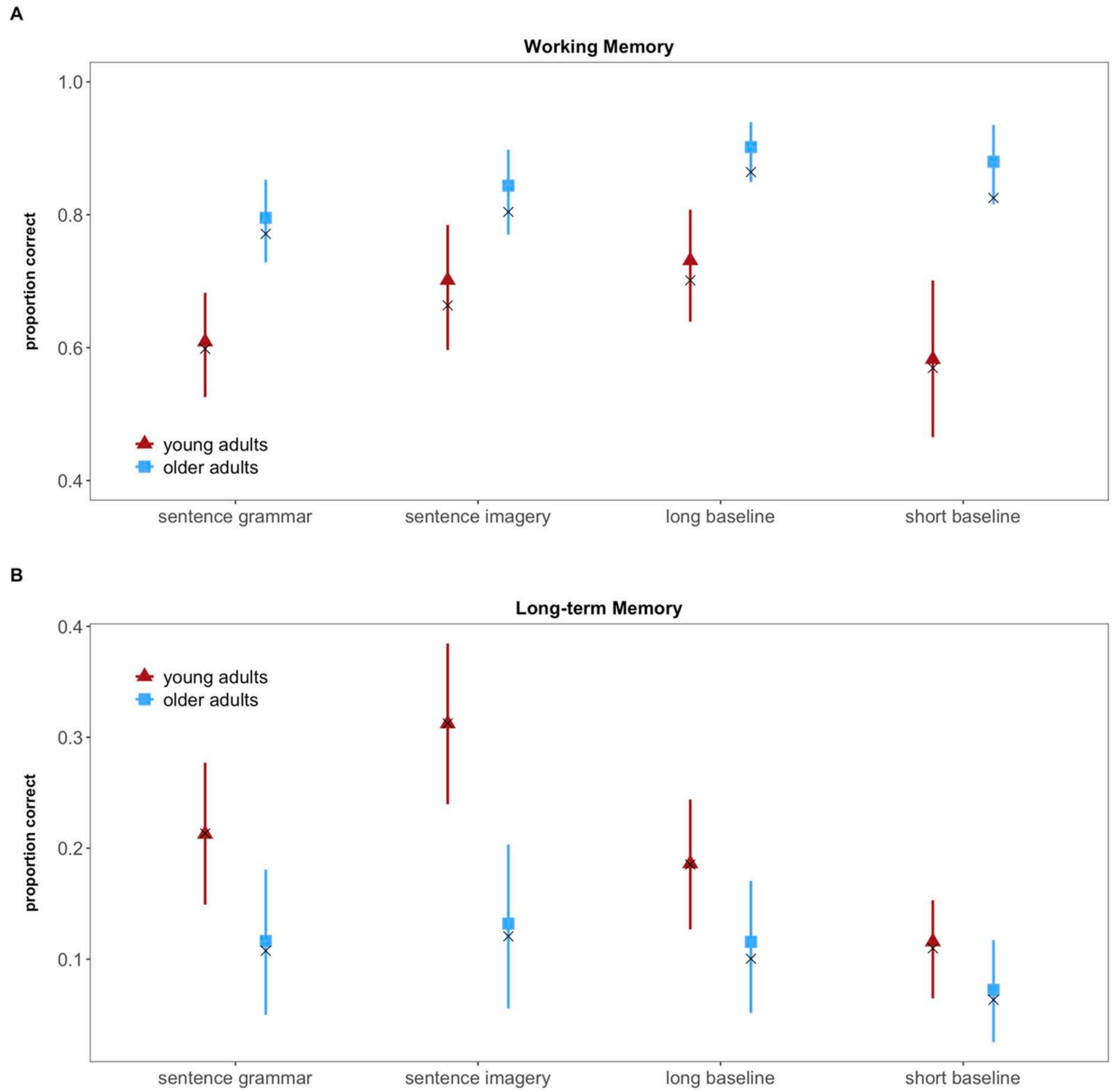

Fig. 5. Proportion correct in the (A) working-memory task and (B) long-term memory task in Experiment 3. The red (young adults) and blue (older adults) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMMs. The crosses represent the observed proportions. Their overlap indicates that the models adequately describe the data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

strategies. In contrast, longer free time led to better LTM only for those trials for which they reported elaborative strategies. Hence, longer free time does increase the opportunity for elaboration, and participants make use of that opportunity, but elaboration improves only episodic LTM, not WM.

The WM results are better explained by the following interpretation: The sentence imagery condition enabled young adults to create durable representations, with deeper associations within LTM, which led to better memory for those words in the LTM test. In line with previous research showing that subjects can flexibly use their LTM in WM tasks (e.g. Lewis-Peacock, Drysdale, & Postle, 2015; Oberauer, Awh, & Sutterer, 2017; Thalmann, Souza, & Oberauer, 2019; Lewis-Peacock and Postle, 2008), they could also draw on these deeper associations during the WM test, thereby improving performance to a level approximating that of the long-baseline condition. However, in the long baseline another process strengthened WM, a process that did not promote more durable or more accessible LTM representations. In older adults, elaboration is less effective, and therefore their WM performance did not even exceed that of the short baseline.

An alternative explanation is that elaboration did improve both WM and LTM in our study, but that the beneficial effect in WM is counteracted by a secondary-task load. A long line of research has shown that carrying out a secondary task that involves processing additional material during encoding or maintenance impairs immediate memory performance (Chein, Moore, & Conway, 2011; Hale, Myerson, Rhee, Weiss, & Abrams, 1996; Jarrold, Tam, Baddeley, & Harvey, 2010). More precisely, the enriched representations created by reading sentences and forming mental images could have created interference in WM in the same way as it is the case for sentence reading in reading span tasks (Daneman & Carpenter, 1980). Elaboration – whether it is experimentally induced or initiated spontaneously – imposes a secondary task demand, and immediate serial recall is known to be vulnerable to secondary tasks (e.g. Jonker & Macleod, 2015, see Oberauer et al., 2018 for an overview (benchmark 5.2.)). Older adults' WM has sometimes been found to suffer more than young adults from a secondary-task demand (Rhodes et al., 2019), and therefore could have been more strongly affected by the secondary task demand of elaboration, thereby showing poorer WM performance in the sentence imagery compared to the long

Table 6
The posterior effect estimates of the pairwise contrasts and their 95% HDRs of the generalized linear mixed model for the immediate serial memory data of Experiment 2.

contrast	age group	Mode of parameter on logit scale	95% HDR
long baseline vs. short baseline	**young**	**0.58**	**[0.18, 0.95]**
	old	**0.49**	**[0.15, 0.84]**
sentence imagery vs. long baseline	young	−0.27	[−0.73, 0.2]
	old	**−0.56**	**[−0.96, −0.15]**
sentence grammar vs. long baseline	**young**	**−0.64**	**[−1.05, −0.25]**
	old	**−0.52**	**[−0.88, −0.16]**
sentence grammar vs. sentence imagery	young	−0.39	[−0.78, 0.04]
	old	0.07	[−0.37, 0.41]
sentence grammar vs. short baseline	young	−0.07	[−0.46, 0.32]
	old	−0.02	[−0.39, 0.32]
sentence imagery vs. short baseline	young	0.3	[−0.2, 0.79]
	old	−0.06	[−0.54, 0.4]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 7
The posterior effect estimates of the pairwise contrasts and their 95% HDRs of the generalized linear mixed model for the delayed memory data of Experiment 2.

contrast	age group	Mode of parameter on proportion-correct scale	95% HDR
long baseline vs. short baseline	**young**	**0.09**	**[0.04, 0.14]**
	old	0.02	[−0.04, 0.07]
sentence imagery vs. long baseline	**young**	**0.17**	**[0.1, 0.24]**
	old	0.03	[−0.04, 0.11]
sentence grammar vs. long baseline	young	0.04	[−0.04, 0.11]
	old	0.02	[−0.06, 0.09]
sentence grammar vs. sentence imagery	**young**	**−0.14**	**[−0.2, −0.07]**
	old	−0.02	[−0.09, 0.05]
sentence grammar vs. short baseline	**young**	**0.13**	**[0.06, 0.18]**
	old	0.03	[−0.03, 0.1]
sentence imagery vs. short baseline	**young**	**0.26**	**[0.18, 0.34]**
	old	0.05	[−0.03, 0.13]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

baseline. This alternative explanation assumes that elaboration has a beneficial effect on WM representations, but the conclusion with regard to the question we started from remains unchanged: Elaboration has no beneficial *net* effect for WM, and therefore the free-time benefit cannot be explained as an effect of elaboration.

The role of mental imagery for elaboration

Our results show that the semantic context provided by the sentences alone had only a modest beneficial effect on LTM, and no credible effect at all on WM. Only the additional instruction to form a mental image resulted in WM performance approximating that of the long baseline, and also promoted the largest LTM effect. Following the dual coding theory, people store associations between two types of information, verbal and visual, separately in LTM (Paivio, 1991). In this way, adding mental imagery to sentence reading results in more retrieval cues than sentence reading alone, which could promote also better immediate recall in a WM task. In line with this claim, a recent review put forward – based on evidence from the imagery literature – that people can use at least two forms of mental representations, verbal as well as depictive representations (Pearson & Kosslyn, 2015). These authors further argue that images contain much implicit information, which makes such depictive representations especially useful for memory. Indeed, evidence suggests an overlap of mental imagery and visual working

Table 8
The posterior effect estimates of the pairwise contrasts and their 95% HDRs of the generalized linear mixed model for the immediate serial memory data of Experiment 3.

contrast	age group	Mode of parameter on logit scale	95% HDR
long baseline vs. short baseline	**young**	**0.64**	**[0.24, 1.03]**
	old	0.19	[−0.35, 0.7]
sentence imagery vs. long baseline	young	−0.16	[−0.55, 0.24]
	old	**−0.53**	**[−1.02, −0.05]**
sentence grammar vs. long baseline	**young**	**−0.55**	**[−0.92, −0.17]**
	old	**−0.82**	**[−1.29, −0.36]**
sentence grammar vs. sentence imagery	**young**	**−0.38**	**[−0.77, −0.01]**
	old	−0.29	[−0.74, 0.15]
sentence grammar vs. short baseline	young	0.1	[−0.31, 0.49]
	old	**−0.63**	**[−1.14, −0.16]**
sentence imagery vs. short baseline	young	0.48	[−0.05, 1.04]
	old	−0.35	[−0.97, 0.25]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 9
The posterior effect estimates of the pairwise contrasts and their 95% HDRs from the generalized linear mixed model for the delayed memory data of Experiment 3.

contrast	age group	Mode of parameter on proportion-correct scale	95% HDR
long baseline vs. short baseline	**young**	**0.08**	**[0.02, 0.13]**
	old	0.04	[−0.01, 0.1]
sentence imagery vs. long baseline	**young**	**0.13**	**[0.06, 0.19]**
	old	0.02	[−0.05, 0.08]
sentence grammar vs. long baseline	young	0.03	[−0.03, 0.09]
	old	0.01	[−0.05, 0.07]
sentence grammar vs. sentence imagery	**young**	**−0.1**	**[−0.18, −0.02]**
	old	−0.01	[−0.09, 0.06]
sentence grammar vs. short baseline	**young**	**0.11**	**[0.04, 0.16]**
	old	0.04	[−0.01, 0.1]
sentence imagery vs. short baseline	**young**	**0.2**	**[0.14, 0.27]**
	old	0.06	[−0.01, 0.12]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

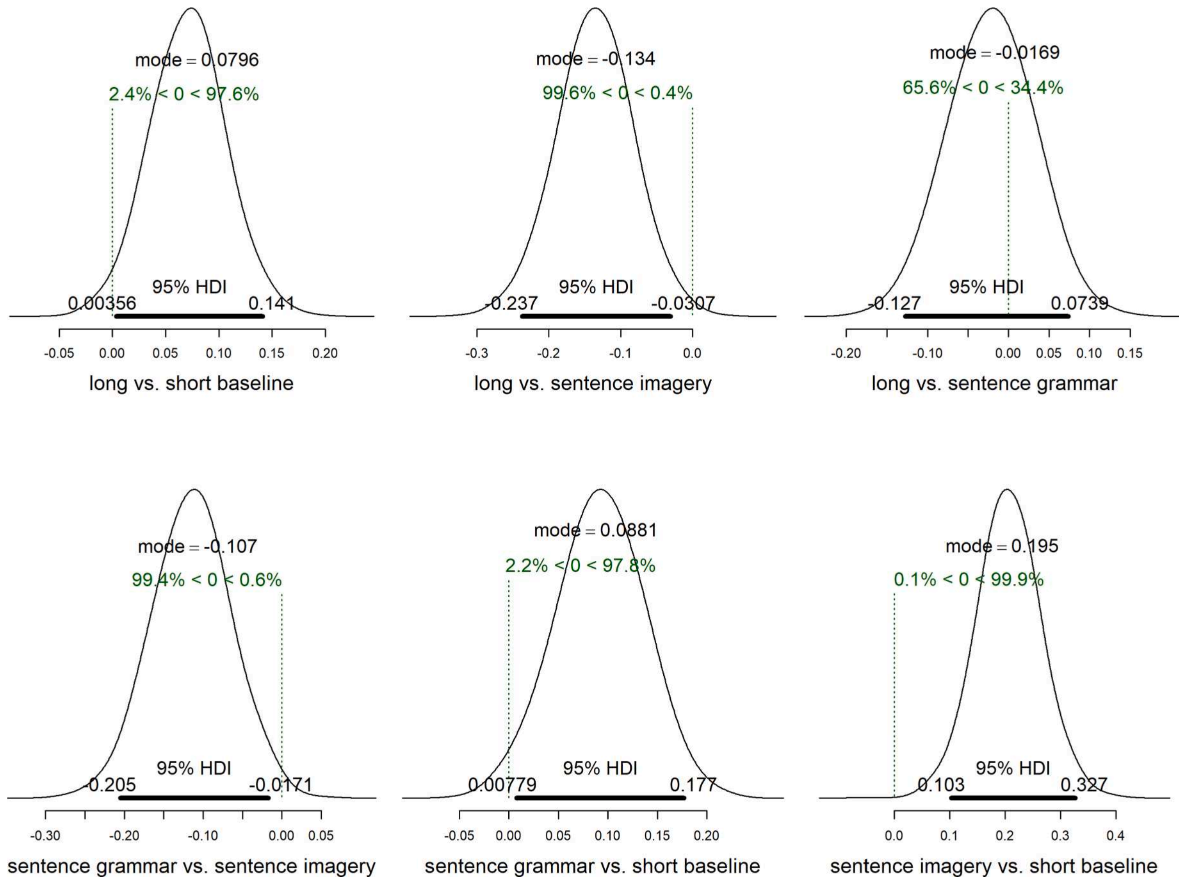

Fig. 6. Posterior distributions of differences between the age groups in the effect of the respective conditions on LTM in Experiment 2. The mode and the highest density intervals reflect the effect size of the differences of pairwise condition contrasts between the age groups s. The dotted line indicates the point of no difference. HDR's including zero reflect that there is no credible interaction of age with the condition contrast.

memory: both share neural correlates and mechanisms in the sensory cortex (e.g. Albers, Kok, Toni, Dijkerman, & De Lange, 2013; see Pearson, Naselaris, Holmes, & Kosslyn, 2015 for an overview), and individuals with higher sensory strength of mental images have been shown to rely more on imagery as a mnemonic strategy in visual WM tasks (Borst, Ganis, Thompson, & Kosslyn, 2012). The present results extend the evidence for the possible beneficial role of mental imagery to a verbal memory task.

Taken together, assisted elaboration through sentences and mental imagery instruction resulted in a larger memory benefit in both immediate and delayed tests compared to merely integrating the to-be-remembered words into a meaningful sentence, attesting to the important role of mental imagery for elaboration.

Older adults' LTM deficit relates to elaboration

As indicated by previous research, older adults present a specific deficit in the effectiveness of elaboration that has been argued to contribute to their decline in LTM. Here, older adults were assisted in two ways: First, elaboration was assisted by the sentences provided, thereby controlling for any deficit in *generating* richer representations and/or associations on the spot. Second, by reducing the memory load at encoding to account for any initial deficits in Experiment 3, we made it easier for older adults to fully process and mentally integrate the sentences. Nevertheless, across both experiments, older adults' episodic LTM did not benefit from either sentence condition compared to the baselines, in contrast to the young adults. The lack of a benefit of the sentence imagery compared to the sentence grammar condition is in line with previous work revealing a reduced benefit of mental imagery strategy use with aging (e.g., Kemps & Newson, 2005; Palladino & De Beni, 2003). Our finding of no benefit of assisted elaboration in old adults contradicts the generation-deficit hypothesis (Smith, 1980), and also rules out the assumption that WM capacity constraints are responsible for older adults' ineffective elaboration.

Our findings disagree with some previous evidence in the field (e.g. Rankin & Collins, 1985) showing that older adults can benefit from elaboration, if richer representations are provided. In contrast to having embedded all to-be-remembered words of a list within one sentence in the present study, Rankin and Collins provided individual sentences per target word. It could be that this provides a more effective form of elaboration for old adults, though we cannot think of a reason why that should be the case.

Another explanation for the lack of an elaboration benefit in older adults could be, that they would have needed more time to process the sentences and to form a mental image. Past research on another form of LTM (associative memory) has shown that time to process information at encoding improves LTM of old adults (Bartsch, Loaiza, & Oberauer, 2019). That said, as subjects adjusted the presentation time at the beginning of the experiment to their personal reading speed, we adjusted for individual differences in speed of sentence processing. Furthermore, in a recent study by Hinault and colleagues, older adults were given 8 s to encode word pairs and form an interactive mental image (compared to 6 s in young adults), yet the beneficial effect of elaboration – compared to rote rehearsal – was much larger in young than older adults (Hinault, Lemaire, & Touron, 2017).

Yet another explanation for the lack of an elaboration benefit in older adults could be, that the extra material in the sentence conditions might have distracted the older adults more than the younger, thereby counteracting any beneficial effect on LTM. Research on the inhibitory deficit hypothesis of aging implemented a similar task – the Reading with

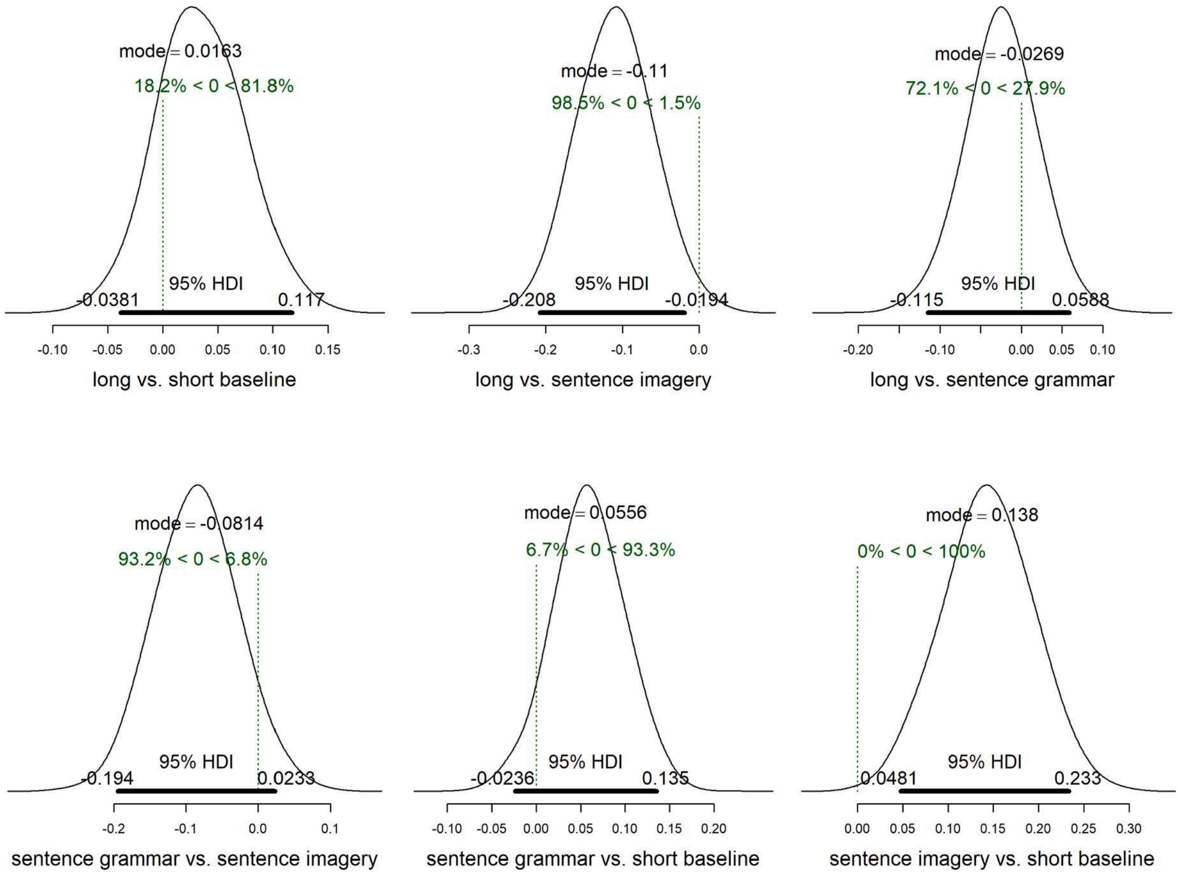

Fig. 7. Posterior distributions of differences between the age groups in the effect of the respective conditions on LTM in Experiment 3. The mode and the highest density intervals reflect the effect size of differences of pairwise condition contrasts between the age groups. The dotted line indicates the point of no difference. HDR's including zero reflect that there is no credible interaction of age with the condition contrast.

Distraction Task – in which subjects must ignore irrelevant words and phrases in order to correctly read the target text of a paragraph. In that task, older compared to young adults typically show increased reading times and errors on comprehension tests (Connelly, Hasher, & Zacks, 1991; Darowski, Helder, Zacks, Hasher, & Hambrick, 2008; Mund, Bell, & Buchner, 2010).[4] However, our task was different from the Reading with Distraction paradigm insofar as all the words presented to the participants were relevant to understanding and reading the sentence. Therefore, our task did not challenge older adults' ability to inhibit distracting information. In support of this assumption, older adults' reading times were only moderately slower than those of young adults (an increase of about 20%, in line with typical age differences in natural-reading speed (Brysbaert, 2019; Liu, Patel, & Kwon, 2017), but smaller than the age difference in reading speed in the Reading with Distraction task).

To conclude, our study confirmed that older adults' LTM benefits very little, if at all, from elaboration. The lack of an elaboration benefit in older adults is not due to an inability to generate the enriched representations needed for elaboration. An earlier study had shown that mental imagery instructions resulted in differentiable brain activation patterns in older adults compared to repeated reading and refreshing – demonstrating that old adults followed the elaboration instruction but, in contrast to younger adults, this did not result in a LTM benefit (Bartsch, Loaiza, Jäncke, et al., 2019). Here we show that even when elaboration is assisted, older adults don't benefit from it. Taken together, older adults are capable to and do generate enriched representations, yet these fail to improve accessibility of episodic memory traces in older adults.

Conclusion

The present study showed that elaboration through embedding words in sentences, and encouraging mental imagery, improves LTM but not WM. Elaboration is not underlying the beneficial effect of free time on working memory. Furthermore, enriched representations fail to improve accessibility of episodic memory traces in older adults, contributing to the pronounced age-related LTM deficit.

CRediT authorship contribution statement

Lea M. Bartsch: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft. **Klaus Oberauer:** Conceptualization, Methodology, Writing - review & editing.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2020.104215.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Psychology of learning and motivation (Vol. 2, pp. 89–195). https://doi.org/10.1016/S0079-7421(08)60422-3.

[4] We thank an anonymous reviewer for pointing us to that literature.

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging, 24*(1), 82–92. https://doi.org/10.1037/a0014078.

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36*(8), 1383–1390. https://doi.org/10.3758/MC.36.8.1383.

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition, 39*(3), 447–461. https://doi.org/10.3758/s13421-010-0034-3.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *NeuroImage, 199*, 585–597. https://doi.org/10.1016/j.neuroimage.2019.06.028.

Bartsch, L. M., Loaiza, V. M., & Oberauer, K. (2019). Does limited working memory capacity underlie age differences in associative long-term memory? *Psychology and Aging, 34*(2), 268–281. https://doi.org/10.1037/pag0000317.

Bartsch, L. M., & Oberauer, K. (2019). Adult age differences in refreshing and elaboration and their consequences for working memory and long-term memory. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/Q76FY.

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition, 46*(5), 796–808. https://doi.org/10.3758/s13421-018-0805-9.

Borst, G., Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2012). Representations in mental imagery and working memory: Evidence from different types of visual masks. *Memory & Cognition, 40*(2), 204–217. https://doi.org/10.3758/s13421-011-0143-7.

Bower, G. H., & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science, 20*(2), 119–120. https://doi.org/10.3758/BF03335632.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language, 109*(April). https://doi.org/10.1016/j.jml.2019.104047.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal, 10*(1), 395–411. https://doi.org/10.32614/rj-2018-017.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software, 76*(1). https://doi.org/10.18637/jss.v076.i01.

Cherry, K. E., Park, D. C., Frieske, D. A., & Rowley, R. L. (1993). The effect of verbal elaborations on memory in young and older adults. *Memory & Cognition, 21*(6), 725–738. https://doi.org/10.3758/BF03202741.

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *Neuroimage, 54*(1), 550–559.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging, 6*(4), 533–541. https://doi.org/10.1037/0882-7974.6.4.533.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*(07), 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in eisodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language, 19*(4), 450–466.

Darowski, E. S., Helder, E., Zacks, R. T., Hasher, L., & Hambrick, D. Z. (2008). Age-related differences in cognition: The role of distraction control. *Neuropsychology, 22*(5), 638–644. https://doi.org/10.1037/0894-4105.22.5.638.

Davachi, L., Maril, A., & Wagner, A. D. (2001). When keeping in mind supports later bringing to mind: Neural markers of phonological rehearsal predict subsequent remembering. *Journal of Cognitive Neuroscience, 13*(8), 1059–1070. https://doi.org/10.1162/089892901753294356.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*(4), 597.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29*(2), 247–253. https://doi.org/10.3758/BF03194918.

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology, 60*(9), 1227–1245. https://doi.org/10.1080/17470210600926075.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6.

Galli, G. (2014). What makes deeply encoded items memorable? Insights into the levels of processing framework from neuroimaging and neuromodulation. *Frontiers in Psychiatry, 5*(MAY), 1–8. https://doi.org/10.3389/fpsyt.2014.00061.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58*(4), 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC.

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin, 102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403.

Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology, 10*(2), 228–240.

Hargreaves, I. S., Pexman, P. M., Johnson, J. C., & Zdrazilova, L. (2012). Richer concepts are better remembered: Number of features effects in free recall. *Frontiers in Human Neuroscience, 6*(April), 1–11. https://doi.org/10.3389/fnhum.2012.00073.

Hinault, T., Lemaire, P., & Touron, D. (2017). Aging effects in sequential modulations of poorer-strategy effects during execution of memory strategies. *Memory, 25*(2), 176–186. https://doi.org/10.1080/09658211.2016.1146300.

Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2010). The nature and position of processing determines why forgetting occurs in working memory tasks. *Psychonomic Bulletin & Review, 17*(6), 772–777.

Jonker, T. R., & Macleod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 831–840. https://doi.org/10.1037/xlm0000069.

Kausler, D. H. (1994). *Learning and memory in normal aging.* Academic Press.

Kemps, E., & Newson, R. (2005). Patterns and predictors of adult age differences in mental imagery. *Aging, Neuropsychology, and Cognition, 12*(1), 99–128. https://doi.org/10.1080/13825580590925152.

Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of Personality and Social Psychology, 55*(1), 5–11. https://doi.org/10.1037/0022-3514.55.1.5.

Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B* (5th ed.). Balingen: Spitta Verlag.

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2015). Neural evidence for the flexible control of mental representations. *Cerebral Cortex, 25*(10), 3303–3313. https://doi.org/10.1093/cercor/bhu130.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. 28(35), 8765–8771. https://doi.org/10.1523/JNEUROSCI.1953-08.2008.Temporary.

Liu, R., Patel, B. N., & Kwon, M. Y. (2017). Age-related changes in crowding and reading speed. *Scientific Reports, 7*(1), 1–10. https://doi.org/10.1038/s41598-017-08652-0.

Loaiza, V. M., & McCabe, D. P. (2013). The influence of aging on attentional refreshing and articulatory rehearsal during working memory on later episodic memory performance. *Aging, Neuropsychology, and Cognition, 20*(4), 471–493. https://doi.org/10.1080/13825585.2012.738289.

McDaniel, M. A., Einstein, G. O., & Jacoby, L. L. (2008). New Considerations in Aging and Memory: The Glass May Be Half Full. In F. I. M. Craik, & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (pp. 251–310). New York and Hove: Psychology Press.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition, 29*(1), 10–17. https://doi.org/10.3758/BF03195736.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8.

Mund, I., Bell, R., & Buchner, A. (2010). Age differences in reading with distraction: Sensory or inhibitory deficits? *Psychology and Aging, 25*(4), 886–897. https://doi.org/10.1037/a0019508.

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning Memory and Cognition, 43*(1), 1–22. https://doi.org/10.1037/xlm0000302.

Oberauer, K., & Lewandowsky, S. (2016). Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm. *Cognition, 156*, 106–128.

Oberauer, K., & Lewandowsky, S. (2019). Simple measurement models for complex working-memory tasks. *Psychological Review, 126*(6), 880–932.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., … Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*(9), 885–958. https://doi.org/10.1037/bul0000153.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45*(3), 255–287.

Palladino, P., & De Beni, R. (2003). When mental images are very detailed: Image generation and memory performance as a function of age. *Acta Psychologica, 113*(3), 297–314. https://doi.org/10.1016/S0001-6918(03)00045-3.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299.

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. Proceedings of the National Academy of Sciences of the United States of America, 112(33), 10089–10092. https://doi.org/10.1073/pnas.1504933112.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences, 19*(10), 590–602.

Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin, 82*(1), 68–84.

Petermann, F., & Wechsler, D. (2012). *Wechsler adult intelligence scale - fourth edition (WAIS-IV)*. Frankfurt am Main: Pearson.

Rankin, J. L., & Collins, M. (1985). Adult age differences in memory elaboration. Retrieved from *Journal of Gerontology, 40*(4), 451–458 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4008880.

Rhodes, S., Jaroslawska, A. J., Doherty, J. M., Belletier, C., Naveh-Benjamin, M., Cowan, N., … Logie, R. H. (2019). Storage and processing in working memory:

Assessing dual-task performance and task prioritization across the adult lifespan. *Journal of Experimental Psychology: General, 148*(7), 1204–1227.
Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5*(4), 597–614.
Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General, 146*(11), 1551–1573. https://doi.org/10.1037/xge0000346.
Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420. https://doi.org/10.1093/beheco/arn145.
Smith, A. D. (1980). Age differences in encoding, storage, and retrieval. In *New directions in memory and aging*. Hillsdale, NJ: Erlbaum.
Shing, Y. L., Werkle-Bergner, M., Brehmer, Y., Müller, V., Li, S. C., & Lindenberger, U. (2010). Episodic memory across the lifespan: The contributions of associative and strategic components. *Neuroscience and Biobehavioral Reviews, 34*(7), 1080–1091. https://doi.org/10.1016/j.neubiorev.2009.11.002.
Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology, 107*, 1–21. https://doi.org/10.1016/j.cogpsych.2018.09.002.
Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15*(3), 535–542.
Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 37–55.
Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*(4), 446–468. https://doi.org/10.1016/S0749-596X(03)00095-0.
Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. *The Handbook of Aging and Cognition, 293–357*. https://doi.org/10.3758/s13423-011-0086-9.